Science
AAAS

www.sciencemag.org/cgi/content/full/321/5897/1781/DC1

Supporting Online Material for

School Performance Will Fail to Meet Legislated Benchmarks

M. J. Bryant, K. A. Hammond, K. M. Bocian, M. F. Rettig, C. A. Miller, R. A. Cardullo*

*To whom correspondence should be addressed. E-mail: cardullo@ucr.edu

Published 26 September 2008, *Science* **320**, 1781 (2008)
DOI: 10.1126/science.1161033

This PDF file includes

Materials and Methods
Figs. S1 and S2
Tables S1 and S2
References

Supporting Online Material

Methods

Multilevel Model Parameterization and Estimation

We modeled student achievement gains using multilevel models. The purpose of these models was to generate estimates of annual growth rates in proficiency, which could then be compared with the legislated increases in the benchmark proficiency levels for meeting AYP. Observed annual growth rates need to be as high or higher than legislated increases in AYP targets in order for all schools to meet the ultimate goal of 100% proficiency by the 2014 deadline.

Data were obtained from public domain research files downloaded from the California Department of Education (*S1*). Specifically, we used the phase I AYP file for 2003 and the annual AYP files from years 2004 to 2007. We only included elementary schools that reported proficiency data for at least four of the five years between 2003 and 2007. To extract data only from elementary schools, we cross-listed the AYP files with the State's API files, which include a variable indicating if the record came from elementary schools. We used the API growth files for 2003, 2004, 2005, 2006, and 2007 solely for the purpose of extracting the elementary school indicator variable. All files were downloaded on 14 October 2007. For the purposes of this study, we did not consider test participation rate or API scores as criteria for meeting AYP.

The level-1 model focuses on changes in achievement over time within a school. We set time equal to zero in 2003 (2004 = 1, 2005 = 2, etc.) to facilitate interpretation of the intercept as the baseline level of achievement. This corresponds to when California's content standards and assessment were fully aligned and the beginning of AYP as an accountability criterion. The level-2 model characterizes the average changes among schools. We estimated an unconditional means model (*S2*) that assumes no changes in achievement, in addition to three different growth models (see below). Model selection was conducted by comparing Akaike information criterion (AIC) values, where the model with the lowest AIC was considered the best fit. Model parameters were estimated using Proc Mixed for the unconditional means and linear and polynomial growth models and Proc NLMixed for the logistic models (*S3*). The fixed effects, their standard errors, and AIC values of the models are reported in Table S2.

The first model estimated was an unconditional means two-level model with the basic form:

1A) Level 1 $Y_{ij} = \pi_{0i} + \varepsilon_{ij}$

1B) Level 2 $\pi_{0i} = \beta_0 + \zeta_{0i}$

Where Y_{ij} is the percent proficient or advanced for the i^{th} school during year j. π_{0i} is the mean achievement for the i^{th} school with residual error ε_{ij}. β_0 is the mean achievement among schools with residual error ζ_{0i}. The model does not take any achievement growth into account by simply treating any year-to-year variation as random. All three growth models have lower AIC values than the unconditional means model, which indicates that achievement does change over time.

The three growth models are extensions of the unconditional means model. The linear growth model produces two fixed-effects terms. β_{00} is the average intercept among schools.

This can be interpreted as a baseline level of achievement or average percent proficiency in 2003. The second coefficient, β_{10} is the average annual rate of achievement growth among schools. The basic equations for the multilevel linear growth are as follows:

2A) Level 1 $Y_{ij} = \pi_{0i} + \pi_{1i}(\text{year}) + \varepsilon_{ij}$

2B1) Level 2 $\pi_{0i} = \beta_{00} + \zeta_{0i}$

2B2) Level 2 $\pi_{1i} = \beta_{10} + \zeta_{1i}$

The multilevel polynomial growth model assumes that level-1 (within schools over time) achievement growth can be described by a polynomial regression. Polynomial models designate parabolas that may be either concave up or concave down. Concave up parabolas can be interpreted as showing accelerated growth as time progresses; concave down parabolas can be interpreted as showing a deceleration in achievement and even a loss of achievement in later years. The basic multilevel polynomial equations are

3A) Level 1 $Y_{ij} = \pi_{0i} + \pi_{1i}(\text{year}) + \pi_{2i}(\text{year}_2) + \varepsilon_{ij}$

3B1) Level 2 $\pi_{0i} = \beta_{00} + \zeta_{0i}$

3B2) Level 2 $\pi_{1i} = \beta_{10} + \zeta_{1i}$

3B3) Level 2 $\pi_{2i} = \beta_{20} + \zeta_{2i}$

The parameters β_{00} and β_{10} are simply the intercept and slopes of a linear regression. The parameter β_{20} is the term that determines if the parabola is concave up ($\beta_{20} > 0$) or if the parabola is concave down ($\beta_{20} < 0$).

The final multilevel growth model we estimated was a logistic model. Logistic models are well suited for values that range from 0 to 1 (such as percentage of students scoring proficient or advanced). The logistic equation is characterized by an initial acceleration followed by a deceleration to an asymptote. The asymptote is 100% proficiency in this context. As with the other linear multilevel models. the first level characterizes each school and level-2 model characterizes an average among schools. The multilevel logistic equations are

4A) Level 1 $Y_{ij} = 1/\{1+[\pi_{0i}e^{-(\pi_{1i}*\text{year})}]\} + \varepsilon_{ij}$

4B1) Level 2 $\pi_{0i} = \beta_{00} + \zeta_{0i}$

4B2) Level 2 $\pi_{1i} = \beta_{10} + \zeta_{1i}$

The parameters π_{0i} and β_{00} are related to an intercept (or baseline); however, in logistic models, the higher the value of π_{0i} or β_{00}, the lower the baseline. The parameters π_{1i} and β_{10} characterize how fast the estimated trajectory approaches the upper asymptote, with larger values corresponding to more rapid attainment of the maximum value.

Determination of Lowest Performing Group or Subgroup

Sanctions are levied against entire schools regardless of the subject area or subgroup that may have failed to meet the threshold proficiency level. For this reason, the lowest proficiency level within a school for a given year may have punitive implications beyond basic informational implications that follow from mere assessment. To be able to compare the distribution of achievement in California elementary schools that will most closely relate to accountability legislation with the achievement distributions for all students on the ELA and mathematics, we found the lowest scoring group or subgroup for both the ELA and mathematics exam.

Using the same public-access research files, we found the lowest percentage of students scoring proficient or advanced in either subject. We only considered the percent proficiency of the subgroups considered numerically significant (essentially 100 or more valid test scores or, for smaller schools, if more than 15% of the students are members of the subgroup). For any school, there are a maximum of 10 sub groups: African American, American Indian, Asian, Filipino, Hispanic, Pacific Islander, White, Students with Disabilities, Socioeconomically Disadvantaged, and English Language Learners.

Some schools are held accountable for more subgroups than other schools, and the subgroups are not mutually exclusive. For example, all students fall into one of the seven ethnicity subgroups, but a student may also be in the socioeconomically disadvantaged subgroup. We classified the lowest scoring group (all students) or subgroups into subject (mathematics or ELA) and into three categories (Socioeconomically Disadvantaged, English Language Learners, and a final category composed of any of the ethnicity subgroups and Students with Disabilities). In the rare (<5% of schools) instances of a tie for lowest percent proficiency, we favored the mathematics exam, because it was readily apparent that the overwhelming majority of lowest proficiency within a school was in ELA. In cases where there was a tie within test subject but between categories of subgroups, we broke ties in a manner that favored categorizing schools as having the lowest percent proficiency in the Socioeconomically Disadvantaged subgroup over English Language Learners subgroup over the combination of all other groups or subgroups. This was done to highlight these two subgroups as they were the two predominantly lowest scoring subgroups out of the 10 possible.

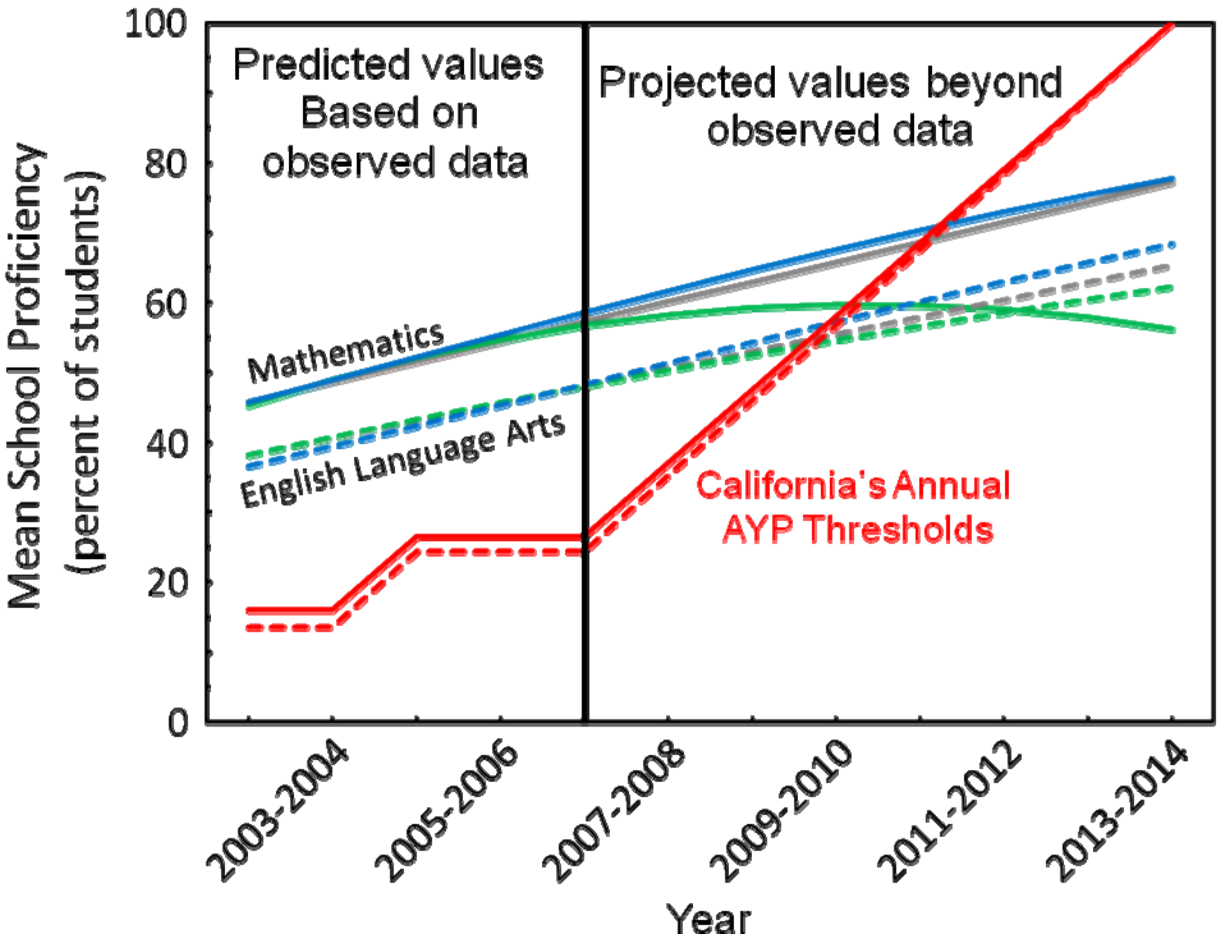

Figure S1. Mean school proficiency compared with AYP benchmarks. Mean school proficiency is defined as the predicted mean number of students scoring proficient or advanced within a school over time from the three multilevel growth models for Mathematics (solid lines) and English Language Arts (dotted lines). The gray lines represent the linear growth model. The green lines represent the polynomial growth model. The blue lines represent the logistic growth model. The red lines indicate California's targeted AYP benchmarks for Mathematics (solid) and English Language Arts (dotted). Parameter estimates for models can be found in Table S2. Models are plotted out to 2014 (beyond the range of available data) to illustrate that the available data (through 2007) does not indicate the accelerated growth in proficiency required to meet legislated goals. The decline in mathematics proficiency (after 2010) illustrated by the polynomial model should be considered an artifact of extrapolation.

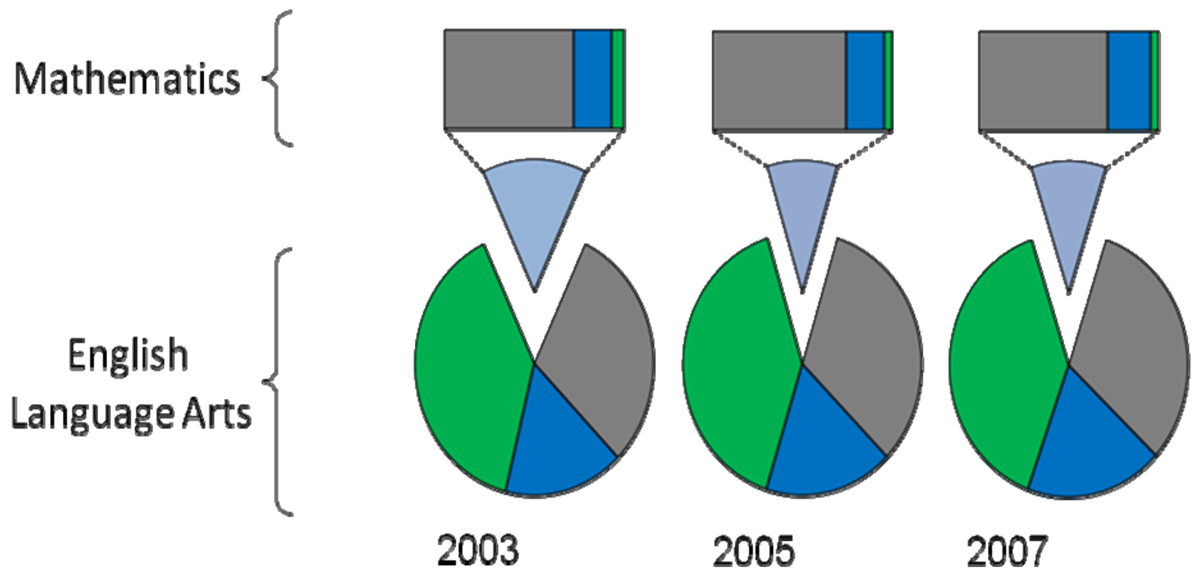

Figure S2. Pie charts of the lowest-scoring subgroups and exam subjects. The green pie slice shows the proportion of schools where the lowest scoring subgroup was English language learners taking the ELA exam. The dark blue pie slice is the proportion of schools where the lowest-scoring subgroup was Socioeconomically Disadvantaged students taking the ELA exam. The gray pie slice is the proportion of schools where the lowest-scoring subgroup comes from the remaining demographic subgroups taking the ELA exam. The light blue pie slice represents the percentage of schools where the lowest-performing subgroup was based on the mathematics exam. The pull-out bar shows the relative proportions by subgroup identity (same color codes are the same as the for the pie slices). These results reveal that it is the English language learners and the Socioeconomically Disadvantaged students' performance on the ELA exam that is most likely initiate sanctions on a school because of failure to meet AYP and that there has been little change in this pattern in the past 5 years. Given this pattern, it is unlikely that these at-risk subgroups will make sufficient progress by 2014.

Table S1. Definitions and consequences of key terms for evaluating educational reform related to the "No Child Left Behind Act of 2001" and California.

Term	Source	Definition	Consequences
Content standards/ statewide assessment	States	Established and mandated expectations of what students at particular grade levels should know and be able to do (in English Language Arts and mathematics). Standardized tests used to determine student proficiency levels in the content standards. For example, California administers the California Standards Test (CST) a criterion referenced test in mathematics, English Language Arts, science, and history.	Because tests are based on the standards and each state determines its own standards, comparisons of student achievement among states is difficult.
Proficiency level/percent proficient	State and Federal	Students are assigned a proficiency level (on an ordinal scale) based on their raw test score. California has five levels: Advanced, Proficient, Basic, Below Basic, and Far Below Basic. The Percent Proficient is the percentage of students within a school who score at least at the proficient level on their statewide assessment.	Schools are held accountable and may be eligible for school improvement funds and may be ultimately sanctioned if they do not meet the statewide benchmarks for percent proficient.
AYP: adequate yearly progress	Federal	A school will either meet or fail to meet AYP. The criteria for meeting AYP include sufficient participation on standardized tests and meeting a benchmark percentage of students achieving proficiency. Benchmarks must be set at 100% proficiency by 2014. However, individual states determine their own rates of growth. Schools are held accountable for test participation and proficiency for all students and all significant ethnic and demographic subgroups.* States may include additional AYP criteria; for example, California has adopted the Academic Performance Index that is an index based on growth in performance on statewide assessments. High schools are also held accountable for graduation rates.	Failure to meet AYP for two consecutive years places a school in Program Improvement sanction

*Recognized subgroups: African American, American Indian, Asian, Filipino, Hispanic, Pacific Islander, White, Students with Disabilities, Socioeconomically Disadvantaged (SD), and English Language Learners (ELL).

Table S2. Estimates for the level-2 (fixed effects) parameters from the multilevel models. For unconditional means, we model average achievement with no possibility for changes in achievement over time. The linear growth model shows linear (constant rate) changes in achievement over time. The polynomial growth model permits acceleration and deceleration in achievement over time. The logistic growth model shows changes in achievement as accelerated growth from a lower followed by decelerated growth towards an upper asymptote. Coefficients are defined in the Methods section. AIC, Akaike information criterion; the model with the lowest AIC is considered to the best model out of the four candidate models. The AIC in bold designates the best model for each test subject.

	Unconditional means		Linear growth		Polynomial growth		Logistic growth	
	ELA	Mathematics	ELA	Mathematics	ELA	Mathematics	ELA	Mathematics
β_{00}: intercept (SE)	0.4318 (0.0027)	0.5157 (0.0024)	0.3822 (0.0028)	0.4584 (0.0026)	0.3814 (0.0028)	0.4529 (0.0026)	1.7320 (0.0231)	1.1901 (0.0142)
β_{10}: slope (SE)	NA	NA	0.0246 (0.0002)	0.0285 (0.0003)	0.0262 (0.0005)	0.0392 (.0007)	0.1203 (0.0012)	0.1294 (0.0015)
β_{20}:polynomial (SE)	NA	NA	NA	NA	–0.0004 (0.0001)	–0.0027 (0.0002)	NA	NA
AIC	–52976	–45444	–69088	–59700	–69270	**–60110**	**–69692**	–60042

References

S1. All data from this study are available for public access from the California Department Educational Accountability system at www.cde.ca.gov/ta/ac/ar/index.asp

S2. J.D Singer, J.B. Willett, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (Oxford Univ. Press, Oxford, 2003).

S3. SAS Institute Inc., SAS 9.1.3 Help and Documentation (SAS Institute Inc., Cary, NC, 2000–2004).